Volume 264, number 1, 109–111 FEBS 08402 May 1990

Cloning and sequence analysis of the *Erythrina corallodendron* lectin cDNA

Rafael Arango[1], Shmuel Rozenblatt[2] and Nathan Sharon[1]

[1] *Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100 and* [2] *Department of Biotechnology, Tel Aviv University, Tel Aviv 9978, Israel*

Received 19 March 1990

Examination of the hemagglutinating activity of extracts from seeds of *Erythrina corallodendron* at various maturation stages revealed that the level of lectin increases markedly past mid-maturation. Seeds at this stage of maturation served as a source of mRNA for the construction of an expression cDNA library in the vector λ Zap, which generates fusion proteins with an N-terminal portion of β-galactosidase. The library was screened with rabbit polyclonal anti-ECorL antiserum. Four immunopositive clones were isolated. Western blot analysis of cell extracts from one of the clones (pIEcL-B) showed a 36 kDa protein that reacted with the antiserum, as well as with a mouse monoclonal antibody raised against the lectin. DNA sequence analysis by the chain termination method revealed that clone pIEcl-C has an insert of 1017 bp with the entire coding sequence of ECorL, beginning with an initiation codon ATG at position 26 and ending with stop codon TAA at position 868. This fragment encodes a polypeptide of 281 amino acids consisting of a signal leader sequence of 25 amino acids and a mature protein of 256 amino acids. The deduced amino acid sequence from this fragment is identical to the sequence of the first 244 amino acids of ECorL, as determined at the protein level, except at 7 positions.

Legume lectin

1. INTRODUCTION

The primary sequence of the first 244 amino acids of the *Erythrina corallodendron* lectin (ECorL) has been recently established in our laboratory by chemical means and found to share extensive homology with other legume lectins [1]. Here we report the isolation and characterization of a cDNA clone representing the entire coding sequence of ECorL. The availability of this clone, in conjunction with the high-resolution X-ray crystallography data of the lectin [2] and its complexes with lactose [3], will enable us to do future studies of the combining site of the lectin by site-directed mutagenesis. Such studies are important for a better understanding of the structural basis of protein/carbohydrate interactions as well as protein/ligand interactions in general. In the long run, they may lead to the production of tailor-made lectins with desired sugar specificities.

2. MATERIALS AND METHODS

2.1. *Materials*

Erythrina corallodendron seeds at different stages of maturation were collected during the months of March–May from trees growing on the Weizmann Institute campus. Restriction enzymes and DNA modifying enzymes were obtained from Amersham, Pharmacia and New England Biolabs. DNA sequencing reagents were obtained from United States Biochemical Corp. λ Zap was purchased from Stratagene.

Correspondence address: R. Arango, Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

2.2. *Purification of poly A⁺ RNA and cDNA cloning*

Total RNA was extracted from seeds past midmaturation (average weight between 0.2 and 0.3 g) by the guanidinium thiocyanate-CsCl method [4]. Poly A^+ RNA was obtained by oligo d(T) cellulose chromatography [5]. Double-stranded cDNA was prepared with Amersham's cDNA synthesis system according to their protocol. The products were blunt-ended with T4 DNA polymerase and methylated with *Eco*RI methylase, followed by addition of *Eco*RI linkers. The linked cDNA was ligated into the expression vector λ Zap [6] and packaged with Amersham's in vitro packaging extracts [7]. The resulting library was then screened without amplification. *E. coli* XL-1 blue cells were used as host.

2.3. *Isolation of lectin cDNA clones*

The screening was done with rabbit polyclonal anti-ECorL antiserum using Stratagene's picoblue immunoscreening kit. Briefly, the recombinant phages with host cells were plated on LB agar plates and incubated at 42°C for 3 h. Nitrocellulose filters previously soaked with isopropyl-β-D-thiogalactopyranoside (IPTG) were placed over the growing plaques and incubated for at least 5 h at 37°C. Replica filters were then placed over the plaques for 12 h at 37°C. The filters were washed and immersed in blocking solution followed by the immunodetection procedure. Rabbit IgG anti-ECorL was detected by alkaline phosphatase conjugated goat anti-rabbit antibody. The system was developed with 5-bromo-4-chloro-indole-phosphate and nitro blue tetrazolium.

Positive clones were purified by several rounds of plating and screening until all plaques in the plates were positive. All the positive clones were rescued from the phage as Bluescript plasmids by coinfection of XL-1 blue cells with the individual λ Zap isolates and VCMSM13 helper phage in accordance with the supplier's instruc-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

tions. The recombinant plasmids were purified and characterized in different ways (see below).

2.4. *Western blot analysis of positive clones*

Crude lysates from bacteria carrying the Bluescript plasmids were size-fractionated in a 10% SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose. Blots were probed with polyclonal rabbit anti-ECorL antiserum and with a mouse monoclonal antibody raised against the lectin. The polyclonal antibodies were detected and developed as done for the plaque immunoscreening. The monoclonal antibodies were detected with a horseradish peroxidase conjugated anti-mouse antibody and revealed with diaminobenzidine.

2.5. *DNA sequence determination*

DNA sequence analysis was carried out by the dideoxy chain termination method [8] using Sequenase, according to the protocol given by United States Biochemical Corp. ECorL cDNA fragment was cleaved with different restriction enzymes and the resulting smaller fragments were subcloned into PUC-8 followed by sequence analysis from double-stranded templates. To complete the sequence for both DNA strands, 3 synthetic oligonucleotides, complementary to the appropriate strand in positions 354–369, 455–472 and 916–932 were used as sequencing primers.

3. RESULTS AND DISCUSSION

3.1. *Cloning of the ECorL cDNA*

Studies on the regulation of expression of soybean agglutinin and other seed storage proteins have shown that their messages increase in seeds that have passed midmaturation [9]. Prior to the cloning experiments, the amount of ECorL in seeds at different maturation stages was estimated semiquantitatively by immunodiffusion using polyclonal antisera and by hemagglutination assays with human erythrocytes. As expected, the level of lectin present in seeds increased very rapidly after midmaturation (data not shown). We therefore used seeds at this stage of maturation as a source of mRNA for the construction of a cDNA library.

A cDNA library of 300000 independent recombinants was constructed in the vector λ Zap which is capable of generating fusion proteins with the N-terminal portion of β-galactosidase [6]. Screening of the library with a rabbit polyclonal anti-ECorL antiserum revealed 4 immunopositive clones designated pIEcL-A, -B, -C and -D. Restriction analysis and cross-hybridization experiments of the 4 clones indicated that they are derived from a single species of mRNA. The sizes of the clones, as analyzed by agarose gel electrophoresis, varied between 950 to 1020 bp. Western blot analysis of cell extracts from one of the clones (pIEcL-B) showed a 36 kDa protein that immunoreacted with the polyclonal antiserum, as well as with a monoclonal anti-lectin antibody (see Fig. 1). This was the expected size for the lectin, taking into consideration a leader sequence and the additional amino acids of β-galactosidase. The complete nucleotide sequence of both DNA strands of this clone and partial sequences of all other clones were determined. The strategy of sequencing is shown in Fig. 2.

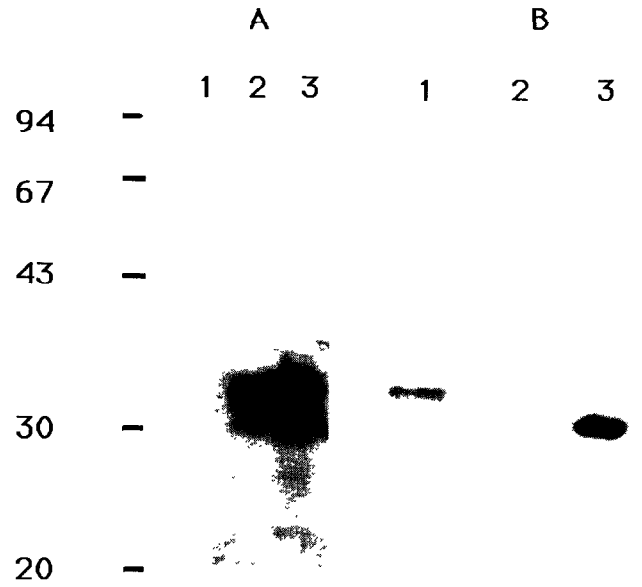

Fig. 1. Western blot analysis of total bacterial lysates of clone pIEcL-B. Western blots were done with rabbit polyclonal anti-ECorL antiserum (panel A) or mouse monoclonal antibodies (panel B). Panel A: (1) Control cells with no vector; (2) clone pIEcL-B grown in the absence of IPTG; (3) clone pIEcL-B grown in the presence of IPTG. Panel B: (1) clone pIEcL-B grown with IPTG; (2) clone pIEcL-B grown without IPTG; (3) ECorL.

As shown in Fig. 3, clone pIEcL-C contains 1017 nucleotides with a coding region of 843 nucleotides beginning with an initiation codon ATG at position 25–27. The fragment also has a 3′ noncoding region of 134 nucleotides before the poly A^+ tail, containing two

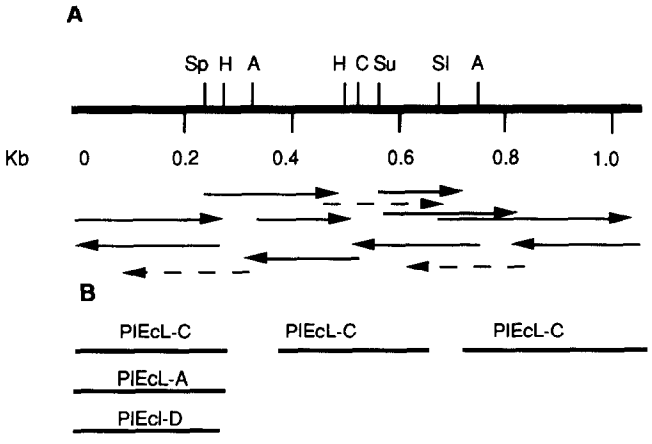

Fig. 2. Sequencing strategy of *E. corallodendron* cDNA. (A) Overlapping restriction fragments of pIEcL-B cDNA were sequenced as indicated by heavy arrows, dashed arrows show regions sequenced with synthetic oligonucleotides (see text). Restriction sites used for the preparation of subclones necessary for sequencing are indicated in the map and are abbreviated as follows: Sp, *Sph*I; H, *Hin*fI; A, *Apa*II; C, *Cla*I; Su, *Sau*3A; Sl, *Sal*I. (B) To complete the ECorL cDNA sequence fragments from clones pIEcL-A, pIEcL-C and pIEcL-D were partially sequenced in the regions indicated by heavy lines.

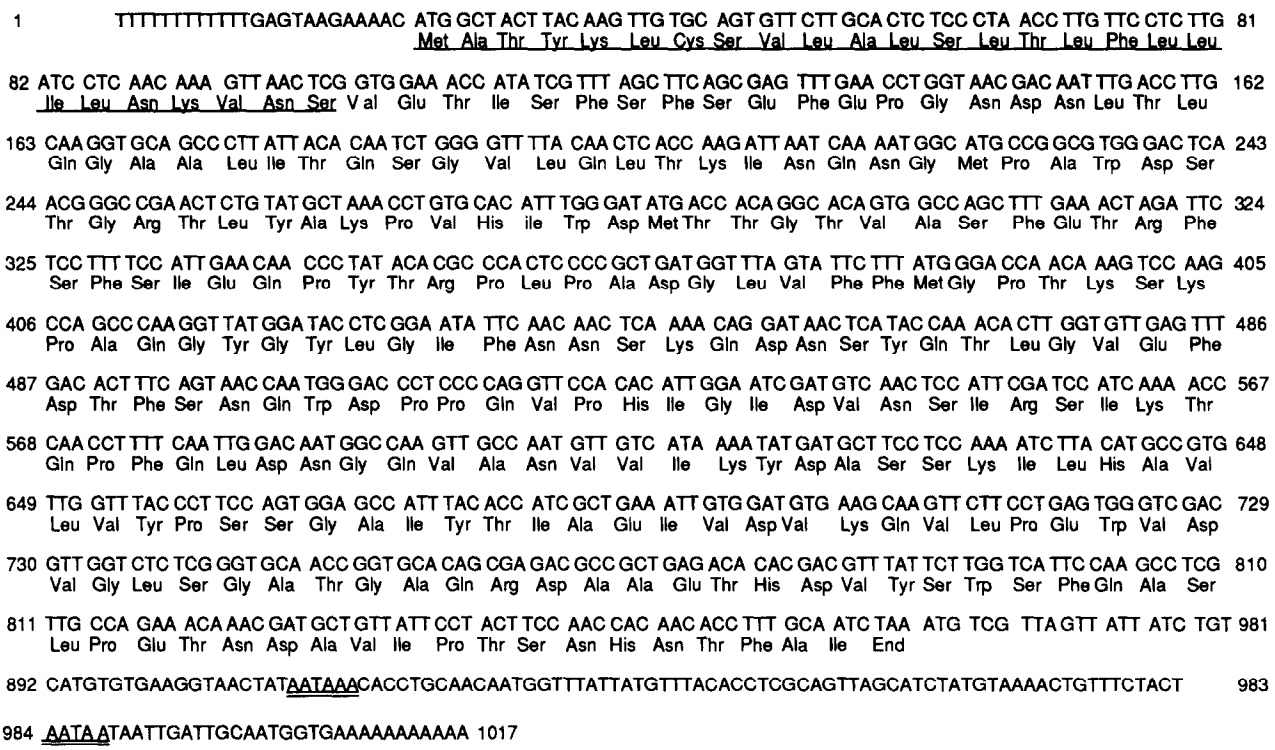

Fig. 3. Sequence of the *Erythrina corallodendron* lectin cDNA and derived amino acid sequence. The solid line indicates the leader sequence of the protein. Two solid lines show polyadenylation signals.

polyadenylation sequences: AATAAA at position 910–916 and AATAAT at position 982–988.

The DNA fragment of pIEcL-C encodes a polypeptide of 281 amino acids with a calculated molecular mass of 30763. Fig. 4 shows that the sequence deduced from pIEcL-C is identical to the sequence of ECorL, as determined by us at the protein level, except at 7 positions: residue 23, Ala instead of Asp; residue 24, Ala instead of Ser; residue 26, Ile instead of Pro; residue 27, Thr instead of Glu; residue 28, Gln instead of Glu; residue 113, Asn instead of Phe and residue 134, Gln instead of Pro. Most of these changes are located in variable regions of the legume lectins except for residue 23 which is Asp in most of them. Asn at position 113 introduces a potential glycosylation site Asn-Asn-Ser.

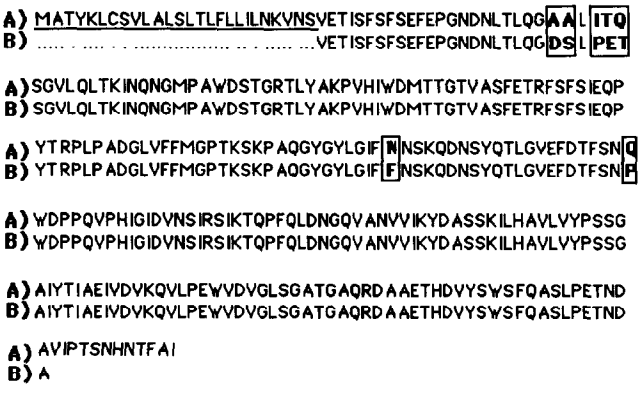

Fig. 4. Comparison of the deduced amino acid sequence of ECorL cDNA (A) to the sequence of the lectin as determined at the protein level (B). The amino acids that are different in the two sequences are boxed. The leader sequence of the lectin is underlined. There are also 12 additional amino acids at the C-terminal of the protein in the deduced sequence.

The sequences of the other positive clones were determined in these regions and were found to be identical to the fragment in pIEcL-C. Thus, the differences between the deduced sequence and the previously determined sequence may be due to the fact that the lectin might be coded by a family of closely related genes. Studies of the genomic organization of other legume lectins have shown the presence of several lectin genes in pea lectin and *Dolichos biflorus* [10,11].

The deduced sequence from pIEcL-C also shows 26 amino acid residues before Val-1 that most likely represent a signal leader peptide needed for translocation of the immature lectin into the endoplasmic reticulum. All legume lectins studied so far have such a signal peptide which is usually between 20 and 30 amino acids long and is characteristically very hydrophobic. At the C-terminal end of the deduced sequence there are 12 additional amino acid residues not found in the chemically determined sequence. This difference may be due to incompleteness of the chemical sequence or alternatively may suggest the possibility of a post-translational modification at the C-terminal of the protein, as has been proposed for other lectins [12].

Comparison of the cDNA sequence of ECorL with the nucleotide sequence of the coding region of soybean agglutinin gave 57% homology and with the cDNA of the *Dolichos biflorus* lectin 50% homology [10,16].

The isolation and characterization of the cDNA of ECorL will allow us to perform experiments aimed at elucidating the structure/function relationships of the lectin at its sugar-binding site. This will be achieved by the expression of the lectin in heterologous systems, as has been done with pea lectin [13,14] and phytohemagglutinin [15], creation of lectin mutants and analysis of the activity of these mutants.

Acknowledgements: Special thanks are due to Dr Halina Lis for her helpful suggestions. Thanks are due to Mrs Segula Haalman and Mrs Hedva Latter for gifts of the lectin and the polyclonal antilectin antibody, respectively, and to Dr David Belinky for the mouse monoclonal antibody. This work was supported by Grant 3727 of the US-Israel Binational Science Foundation.

REFERENCES

[1] Adar, R., Richardson, M., Lis, H. and Sharon, N. (1989) FEBS Lett. 257, 81–85.
[2] Saper, M.A., Lis, H., Sharon, N. and Shaanan, B. (1987) J. Mol. Biol. 193, 823–824.
[3] Shaanan, B., Adar, R., Richardson, M., Lis, H. and Sharon, N. (1989) The Primary Sequence and Three Dimensional Structure of *Erythrina corallodendron* Lectin, International Symposium on Lectins, September 17–19.
[4] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[5] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[6] Short, J.M., Fernandez, J., Sorge, J.M. and Huse, W.D. (1988) Nucleic Acids Res. 16, 7583–7600.
[7] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1988) Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience, New York.
[8] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] Walling, L., Drews, G.N. and Goldberg, B.R. (1986) Proc. Natl. Acad. Sci. USA 83, 2123–2127.
[10] Schnell, D.J. and Etzler, M.E. (1988) J. Biol. Chem. 263, 1468–1453.
[11] Kaminski, P.A., Buffard, D. and Strosberg, D. (1987) Plant Mol. Biol. 9, 497–507.
[12] Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C. and Spencer, D. (1983) J. Biol. Chem. 258, 9544–9549.
[13] Stubbs, M.E., Carver, P.J. and Dunn, R.J. (1986) J. Biol. Chem. 14, 6141–6144.
[14] Prasthofer, T., Phillips, S.R., Suddath, F.L. and Engler, J.A. (1989) J. Biol. Chem. 264, 6793–6796.
[15] Hoffman, L.M. and Donaldson, D.D. (1987) Bio/Technology 5, 157–160.
[16] Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) Cell 34, 1023–1031.